

⑫

DEMANDE DE BREVET D'INVENTION

A1

②2 Date de dépôt : 19.12.01.

③0 Priorité :

④3 Date de mise à la disposition du public de la  
demande : 20.06.03 Bulletin 03/25.

⑤6 Liste des documents cités dans le rapport de  
recherche préliminaire : *Se reporter à la fin du  
présent fascicule*

⑥0 Références à d'autres documents nationaux  
apparentés :

⑦1 Demandeur(s) : THOMSON LICENSING S.A. Société  
anonyme — FR.

⑦2 Inventeur(s) : LE CLERC FRANCOIS et MAREC  
SYLVAIN.

⑦3 Titulaire(s) :

⑦4 Mandataire(s) : THOMSON MULTIMEDIA.

⑤4 PROCÉDE D'ESTIMATION DU MOUVEMENT DOMINANT DANS UNE SEQUENCE D'IMAGES.

⑤7 Le procédé effectuant un calcul (1) d'un champ de  
vecteurs de mouvement associé à une image, définissant,  
pour un élément d'image de coordonnées  $x_i, y_i$ , un ou plu-  
sieurs vecteurs de mouvement de composantes  $u_i, v_i$ , est  
caractérisé en ce qu'il réalise également les étapes  
suivantes:

- modélisation du mouvement (2) à partir d'une repré-  
sentation paramétrique simplifiée:

$$u_i = tx + k.x_i$$

$$v_i = ty + k.y_i$$

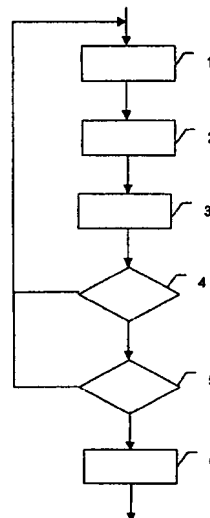
avec

- $tx, ty$  composantes d'un vecteur représentant la compo-  
sante de translation du mouvement,
- $k$  facteur de divergence caractérisant la composante de  
zoom du mouvement,

- régression linéaire robuste (3) dans chacun des deux  
espaces de représentation du mouvement définis par les  
plans  $(x, u)$  et  $(y, v)$ ,  $x, y, u$  et  $v$  représentant respectivement  
les axes des variables  $x_i, y_i, u_i$  et  $v_i$ , pour donner des droites  
de régression,

- calcul des paramètres  $tx, ty$ , et  $k$  (4, 5) à partir des pen-  
tes et des ordonnées à l'origine des droites de régression.

Les applications concernent la sélection d'images-clés  
pour l'indexation vidéo ou la génération de méta-données.



L'invention concerne un procédé et dispositif d'estimation du mouvement dominant dans un plan vidéo. Plus précisément, le procédé se base sur l'analyse des champs de mouvement transmis avec la vidéo dans des schémas de compression utilisant une compensation du mouvement. De tels schémas sont mis en œuvre dans les normes de compression vidéo MPEG-1, MPEG-2 et MPEG-4.

Il est connu des procédés d'analyse de mouvement basés sur l'estimation, à partir des vecteurs de mouvement issus des flux vidéo comprimés de type MPEG, d'un modèle de mouvement qui est le plus souvent affine:

$$\begin{cases} u(x_i, y_i) = ax_i + by_i + c \\ v(x_i, y_i) = dx_i + ey_i + f \end{cases}$$

où  $u$  et  $v$  sont les composantes d'un vecteur  $\vec{\omega}_i$  présent à la position  $(x_i, y_i)$  du champ de mouvement. L'estimation des paramètres affines  $a, b, c, d, e$  et  $f$  du modèle de mouvement s'appuie sur une technique de minimisation des erreurs aux moindres carrés. Un tel procédé est décrit dans l'article de M.A. Smith et T. Kanade « Video Skimming and Characterization through the Combination of Image and Language Understanding » (actes de IEEE 1998 International Workshop on Content-Based Access of Image and Video Databases, pages 61 à 70). Les auteurs de cet article utilisent les paramètres du modèle affine du mouvement, ainsi que les moyennes  $\bar{u}$  et  $\bar{v}$  des composantes spatiales des vecteurs du champ, pour identifier et classifier le mouvement apparent. Par exemple, pour déterminer si le mouvement est un zoom, ils vérifient qu'il existe un point de convergence  $(x_0, y_0)$  du champ de vecteurs, tel que  $u(x_0, y_0) = 0$  et  $v(x_0, y_0) = 0$ , au moyen de la condition suivante :

$$\begin{vmatrix} a & b \\ d & e \end{vmatrix} \neq 0$$

Les moyennes des composantes des vecteurs  $\bar{u}$  et  $\bar{v}$  sont analysées pour tester l'hypothèse d'un panoramique.

Il est également connu des procédés d'analyse de mouvement exploitant directement les champs de vecteurs issus du flux vidéo MPEG, sans passer par l'identification d'un modèle de mouvement. L'article de O.N.

Gerek et Y. Altunbasak « Key Frame Selection from MPEG Video Data » (actes du congrès Visual Communications and Image Processing '97, pages 920 à 925) décrit un tel procédé. La méthode consiste à construire, pour chaque champ de mouvement associé à une image du train binaire MPEG, deux histogrammes du champ de vecteurs, l'un relevant l'occurrence des vecteurs en fonction de leur direction, et le second en fonction de leur amplitude. Des exemples de tels histogrammes sont représentés sur les figures 1 et 2 : la figure 1 illustre une configuration où le mouvement apparent dans l'image est un zoom, alors que dans la figure 2 le mouvement dominant est un panoramique.

Un seuillage de la variance associée au nombre de vecteurs de mouvement dans chaque classe (« bin » en anglais) de l'histogramme, pour chacun des deux histogrammes, est ensuite utilisé pour identifier la présence de mouvements dominants de type "zoom" et "panoramique".

Les méthodes telles que celle proposée par Gerek et Altunbasak fournissent une information purement qualitative sur la catégorie du mouvement dominant, alors qu'une estimation quantitative sur l'amplitude du mouvement est souvent requise. Les méthodes telles que celle proposée par Smith et Kanade basées sur l'estimation d'un modèle paramétrique de mouvement fournissent ces informations quantitatives, mais sont souvent assez peu fiables. En effet, ces méthodes ne prennent pas en compte la présence dans la scène vidéo traitée de plusieurs objets suivant des mouvements apparents différents. La prise en compte des vecteurs associés aux objets secondaires est susceptible de fausser de manière significative l'estimation aux moindres carrés des paramètres du modèle de mouvement dominant. Un objet secondaire est ici défini comme un objet qui occupe sur l'image une surface inférieure à celle d'au moins un autre objet de l'image, l'objet associé au mouvement dominant étant celui qui occupe la plus grande surface dans l'image. Par ailleurs, même en présence d'un seul objet en mouvement dans l'image, les vecteurs du flux vidéo comprimé qui servent de base à l'analyse du mouvement ne reflètent pas toujours la réalité du mouvement apparent réel de l'image. En effet, ces vecteurs ont été calculés dans le but de minimiser la quantité d'information à transmettre après compensation de mouvement, et non d'estimer le mouvement physique des pixels de l'image.

Une estimation fiable d'un modèle de mouvement à partir des vecteurs issus du flux comprimé requiert l'utilisation d'une méthode robuste, éliminant automatiquement du calcul les vecteurs de mouvement relatifs à des objets secondaires ne suivant pas le mouvement dominant, ainsi que les  
5 vecteurs ne correspondant pas au mouvement physique de l'objet principal de l'image.

Des méthodes robustes d'estimation d'un modèle paramétrique de mouvement dominant ont déjà été proposées dans des contextes différents de l'utilisation de flux vidéo comprimés. Un exemple en  
10 est fourni dans l'article de P. Bouthemy, M. Gelgon et F. Ganansia intitulé « A unified approach to shot change detection and camera motion characterization », paru dans la revue IEEE Circuits and Systems for Video Technology volume 9 n° 7, Octobre 1999, pages 1030 à 1044. Ces méthodes présentent l'inconvénient d'être très complexes à mettre en  
15 oeuvre.

L'invention présentée ici a pour but de pallier les inconvénients des différentes familles de méthodes d'estimation du mouvement dominant présentées ci-dessus.

L'invention a pour objet un procédé de détection d'un  
20 mouvement dominant dans une séquence d'images effectuant un calcul d'un champ de vecteurs de mouvement associé à une image, définissant, pour un élément d'image de coordonnées  $x_i$ ,  $y_i$ , un ou plusieurs vecteurs de mouvement de composantes  $u_i$ ,  $v_i$ , caractérisé en ce qu'il effectue également les étapes suivantes:

25 - modélisation du mouvement à partir d'une représentation paramétrique simplifiée :

$$u_i = t_x + k.x_i$$

$$v_i = t_y + k.y_i$$

avec

30  $t_x$ ,  $t_y$  composantes d'un vecteur représentant la composante de translation du mouvement,

$k$  facteur de divergence caractérisant la composante de zoom du mouvement,

- régression linéaire robuste dans chacun des deux espaces de  
35 représentation du mouvement définis par les plans  $(x,u)$  et  $(y,v)$ ,  $x$ ,  $y$ ,  $u$  et  $v$

représentant respectivement les axes des variables  $x_i$ ,  $y_i$ ,  $u_i$  et  $v_i$ , pour donner des droites de régression,

- calcul des paramètres  $t_x$ ,  $t_y$ , et  $k$  à partir des pentes et des ordonnées à l'origine des droites de régression.

- 5                    Selon un mode de mise en œuvre, la régression robuste est la méthode de la moindre médiane des carrés qui consiste à chercher, parmi un ensemble de droites  $j$ ,  $r_{i,j}$  étant le résidu du  $i$ ème échantillon de coordonnées  $x_i$ ,  $u_i$  ou  $y_i$ ,  $v_i$ , par rapport à une droite  $j$ , celle fournissant la valeur médiane de l'ensemble des carrés des résidus qui est minimale:

$$\min_j (\text{med}_i r_{i,j}^2)$$

10

Selon un mode de mise en oeuvre, la recherche de la moindre médiane des carrés des résidus est appliquée sur un nombre prédéfini de droites déterminées chacune par un couple d'échantillons tirés aléatoirement dans l'espace de représentation du mouvement considéré.

- 15                    Selon un mode de mise en œuvre, le procédé effectue, après la régression linéaire robuste, une deuxième régression linéaire non-robuste permettant d'affiner les estimations des paramètres du modèle de mouvement. Cette deuxième régression linéaire peut exclure les points dans les espaces de représentation dont le résidu de régression issu de la
- 20 première régression robuste excède un seuil prédéterminé.

- Selon un mode de mise en œuvre, le procédé effectue un test d'égalité des coefficients directeurs des droites de régression calculées dans chacun des espaces de représentation, basé sur une comparaison des sommes des carrés des résidus obtenus premièrement en effectuant deux
- 25 régressions séparées dans chaque espace de représentation, deuxièmement en effectuant une régression de pente globale sur l'ensemble des échantillons des deux espaces de représentation, et, dans le cas où le test est positif, estime le paramètre  $k$  du modèle par la moyenne arithmétique des coefficients directeurs des droites de régression obtenues dans chaque
- 30 espace de représentation.

L'invention concerne également un dispositif comportant des moyens de calcul pour la mise en œuvre du procédé.

- En exploitant un modèle paramétrique très simplifié, mais
- 35 néanmoins suffisamment réaliste, du mouvement dominant dans une image

vidéo, le procédé permet la mise en oeuvre de méthodes robustes d'identification du modèle de mouvement à un coût réduit. Plus précisément, l'intérêt principal du procédé décrit dans l'invention réside dans l'utilisation d'un espace de représentation judicieux des composantes des vecteurs de mouvement, qui permet de ramener l'identification des paramètres du modèle de mouvement à une double régression linéaire.

D'autres particularités et avantages de l'invention apparaîtront clairement dans la description suivante donnée à titre d'exemple non limitatif et faite en regard des figures annexées qui représentent :

- la figure 1, un champ de vecteurs de mouvement théoriques correspondant à un "zoom",
- la figure 2, un champ de vecteurs de mouvement théoriques correspondant à une scène pour laquelle le mouvement dominant de l'arrière-plan est de type "panoramique", et qui comporte également un objet secondaire suivant un mouvement distinct du mouvement dominant,
- la figure 3, une illustration des espaces de représentation des vecteurs de mouvement utilisés dans l'invention,
- la figure 4, la distribution des vecteurs théoriques pour un mouvement de zoom centré dans les espaces de représentation utilisés dans l'invention,
- la figure 5, la distribution des vecteurs théoriques pour un mouvement de translation oblique globale de l'image dans les espaces de représentation utilisés dans l'invention,
- la figure 6, la distribution des vecteurs théoriques pour un mouvement combiné de translation et zoom dans les espaces de représentation utilisés dans l'invention,
- la figure 7, la distribution des vecteurs théoriques pour une scène statique (mouvement nul) dans les espaces de représentation utilisés dans l'invention,
- la figure 8, un organigramme de la méthode de détection du mouvement dominant.

La caractérisation du mouvement dominant dans une séquence d'images passe par l'identification d'un modèle paramétrique de mouvement dominant apparent. Dans le contexte de l'exploitation de champs de vecteurs

de mouvement issus de flux vidéo comprimés, ce modèle doit représenter le mouvement apparent dans le plan image 2D. Un tel modèle est obtenu par approximation de la projection sur le plan image du mouvement des objets dans l'espace tri-dimensionnel. A titre d'exemple, le modèle affine à 6  
 5 paramètres  $(a,b,c,d,e,f)$  présenté plus haut est couramment adopté dans la littérature.

Le procédé proposé consiste, à la base, à identifier ce modèle paramétrique du mouvement, à partir de champs de vecteurs de mouvement qui sont fournis dans le flux vidéo afin d'en effectuer le décodage, lorsque le  
 10 principe de codage fait appel à des techniques de compensation de mouvement telles qu'exploitées par exemple dans les normes MPEG-1, MPEG-2 et MPEG-4. Toutefois, le procédé décrit dans l'invention est également applicable à des champs de vecteurs de mouvement qui auraient été calculés par un processus séparé à partir des images constituant la  
 15 séquence vidéo traitée.

Dans le contexte de la présente invention, le modèle de mouvement adopté est dérivé d'un modèle linéaire simplifié à 4 paramètres  $(t_x, t_y, k, \theta)$  que nous appellerons SLM (acronyme de l'expression anglaise Simplified Linear Model), défini par :

20

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} = \begin{bmatrix} t_x \\ t_y \end{bmatrix} + \begin{bmatrix} k & -\theta \\ \theta & k \end{bmatrix} \begin{bmatrix} x_i - x_g \\ y_i - y_g \end{bmatrix}$$

avec :

$(u_i, v_i)^t$  : composantes du vecteur de mouvement apparent  
 25 associé au pixel du plan image de coordonnées  $(x_i, y_i)$ ,

$(x_g, y_g)^t$  : coordonnées du point de référence pour l'approximation de la scène 3D filmée par la caméra en une scène 2D ; ce point de référence sera assimilé au point de coordonnées  $(0, 0)^t$  de l'image,

$(t_x, t_y)^t$  : vecteur représentant la composante de translation  
 30 du mouvement,

$k$  : terme de divergence représentant la composante de zoom du mouvement,

$\theta$ : angle de rotation du mouvement autour de l'axe de la caméra.

5 L'objectif recherché est d'identifier les mouvements dominants provoqués par les déplacements et les transformations optiques des caméras, par exemple un zoom optique, dans les séquences vidéo. Il s'agit en particulier d'identifier les mouvements de caméra qui sont statistiquement les plus répandus dans la composition des documents vidéo, regroupant  
10 majoritairement les mouvements de translation et de zoom, leur combinaison, et les absences de mouvement, c'est à dire les plans statiques ou fixes. Les effets de rotation de caméra, très rarement observés en pratique, ne sont pas pris en compte : on restreint donc le modèle aux 3 paramètres  $(t_x, t_y, k)$  en effectuant l'hypothèse que  $\theta \approx 0$ .

15 On a alors deux relations de linéarité entre les composantes des vecteurs et leur position spatiale dans l'image :

$$\begin{cases} u_i = t_x + k.x_i \\ v_i = t_y + k.y_i \end{cases}$$

L'avantage de cette représentation paramétrique simplifiée du mouvement est que les paramètres  $t_x$ ,  $t_y$  et  $k$ , décrivant respectivement les  
20 deux composantes de translation et le paramètre de zoom du modèle de mouvement, peuvent être estimés par régression linéaire dans les espaces de représentation du mouvement  $u_i = f(x_i)$  et  $v_i = f(y_i)$ . Ainsi que l'illustre la figure 3, la représentation d'un champ de vecteurs de mouvement dans ces espaces fournit généralement, pour chacun d'entre eux, un nuage de points  
25 distribué autour d'une droite de pente  $k$ .

Le processus d'estimation des paramètres du modèle simplifié de mouvement est basé sur l'application d'une régression linéaire de type robuste dans chacun des espaces de représentation du mouvement. La régression linéaire est une opération mathématique qui détermine la droite  
30 s'ajustant au mieux à un nuage de points, par exemple en minimisant la somme des carrés des distances de chaque point à cette droite. Cette opération est, dans le contexte de l'invention, mise en œuvre à l'aide d'une technique d'estimation statistique robuste, de façon à garantir une certaine insensibilité vis-à-vis de la présence de valeurs aberrantes dans les



données. En effet, l'estimation du modèle de mouvement dominant doit s'affranchir:

- de la présence dans l'image de plusieurs objets dont certains suivent des mouvements secondaires distincts du mouvement dominant,
- 5       - de la présence de vecteurs de mouvements ne représentant pas le mouvement physique des objets. En effet, les vecteurs de mouvement transmis dans un flux vidéo comprimé ont été calculés dans le but de minimiser la quantité d'information résiduelle à transmettre après compensation de mouvement et non dans le but de fournir le mouvement
- 10   réel des objets constituant la scène imagée.

La figure 8 synthétise les différentes étapes de la méthode d'estimation du mouvement dominant dans la séquence. Chacune de ces étapes est décrite plus précisément dans ce qui suit.

- Une première étape 1 effectue une normalisation des champs
- 15   de vecteurs de mouvement associés chacun à une image de la séquence vidéo traitée. Ces champs de vecteurs sont supposés avoir été calculés préalablement à l'application de l'algorithme, à l'aide d'un estimateur de mouvement. L'estimation du mouvement peut être effectuée pour des blocs rectangulaires de pixels de l'image, comme dans les méthodes
- 20   d'appariement de blocs de pixels dites « block-matching », ou fournir un champ de vecteurs dense, où un vecteur est estimé pour chaque pixel de l'image. La présente invention traite préférentiellement, mais non exclusivement, le cas où les champs de vecteurs utilisés ont été calculés par un encodeur vidéo et transmis dans le flux vidéo comprimé à des fins de
- 25   décodage. Dans le cas typique où le schéma d'encodage utilisé est conforme à une des normes MPEG-1 ou MPEG-2, les vecteurs de mouvement sont estimés pour l'image courante à raison d'un vecteur par bloc rectangulaire de l'image, relativement à une trame de référence dont la distance temporelle à l'image courante est variable. Par ailleurs, pour certaines trames dites « B »
- 30   prédites de manière bi-directionnelle, deux vecteurs de mouvement peuvent avoir été calculés pour un même bloc, l'un pointant depuis l'image courante vers une trame de référence passée et l'autre depuis l'image courante vers une trame de référence future. Une étape de normalisation des champs de vecteurs est donc indispensable de façon à traiter, dans les étapes
- 35   ultérieures, des vecteurs calculés sur des intervalles temporels de durées égales et pointant dans une même direction. Le paragraphe 3.2 de l'article

de V. Kobla et D. Doermann intitulé « Compressed domain video indexing techniques using DCT and motion vector information in MPEG video », Proceedings of the SPIE vol. 3022, 1997, pages 200 à 211, fournit un exemple de méthode permettant d'effectuer cette normalisation. D'autres techniques plus simples basées sur des approximations linéaires du mouvement sur les intervalles de calcul des vecteurs MPEG peuvent également être utilisées.

La deuxième étape référencée 2 effectue une construction des espaces de représentation du mouvement présentés plus haut. Chaque vecteur  $\vec{\omega}_i$  du champ de mouvement, de composantes  $(u_i, v_i)^t$  et de position  $(x_i, y_i)^t$ , est représenté par un point dans chacun des deux espaces  $u_i=f(x_i)$  et  $v_i=f(y_i)$ .

Chaque couple de points  $(x_i, u_i)$  et  $(y_i, v_i)$  correspondant à la représentation d'un vecteur du champ de mouvement peut être modélisé relativement aux droites de régression dans chacun des espaces par :

$$\begin{cases} u_i = a_0 \cdot x_i + b_0 + \varepsilon_{u_i} \\ v_i = a_1 \cdot y_i + b_1 + \varepsilon_{v_i} \end{cases}$$

où

$(a_0, b_0)$  sont les paramètres de la droite de régression à calculer dans l'espace  $u_i=f(x_i)$  ;  $\varepsilon_{u_i}$  est l'erreur résiduelle correspondante.

$(a_1, b_1)$  sont les paramètres de la droite de régression à calculer dans l'espace  $v_i=f(y_i)$  ;  $\varepsilon_{v_i}$  est l'erreur résiduelle correspondante.

La figure 3 illustre des nuages de points obtenus après construction de ces deux espaces à partir d'un champ de vecteurs de mouvement normalisé.

Les paramètres  $(a_0, b_0)$  et  $(a_1, b_1)$  obtenus à l'issue des régressions linéaires dans chacun des espaces de représentation fournissent des estimations des paramètres du modèle de mouvement dominant. Ainsi, les pentes  $a_0$  et  $a_1$  correspondent à une double estimation du paramètre de divergence  $k$  caractérisant la composante de zoom, alors que les ordonnées à l'origine  $b_0$  et  $b_1$  correspondent à une évaluation des composantes  $t_x$  et  $t_y$  de translation.

Les figures 4 à 7 montrent quelques exemples de configurations possibles.

- distribution des données en cas de zoom centré pour la figure 4,

- distribution des données en cas de mouvement de translation oblique pour la figure 5,

5 - distribution des données en cas de zoom décentré (mouvement combinant un zoom et une translation) pour la figure 6,  
 - distribution des données en cas d'une absence de mouvement pour la figure 7.

10 L'étape suivante 3 effectue une régression linéaire robuste pour chacun des espaces de représentation du mouvement, dans le but de séparer les points de données représentatifs du mouvement dominant réel de ceux correspondant, soit au mouvement d'objets secondaires dans l'image, soit à des vecteurs ne traduisant pas le mouvement physique des pixels auxquels ils sont associés.

15 Il existe plusieurs familles de techniques d'estimation robuste. Selon un mode préférentiel de réalisation de l'invention, les droites de régression sont calculées de manière à satisfaire le critère de la moindre médiane des carrés. La méthode de calcul, présentée brièvement ci-après, est décrite de manière plus complète au paragraphe 3 de l'article de P. Meer,  
 20 D. Mintz et A. Rosenfeld « Robust Regression Methods for Computer Vision : A Review », paru dans International Journal of Computer Vision, volume 6 n° 1, 1991, pages 59 à 70.

Appelant  $r_{ij}$  le résidu du  $i^{\text{ème}}$  échantillon d'un espace de représentation du mouvement dans lequel on cherche à estimer l'ensemble  
 25  $E_j$  des paramètres de régression (pente et intercept de la droite de régression), on calcule  $E_j$  de manière à satisfaire le critère suivant :

$$\min_{E_j} (\text{med}_i r_{i,j}^2)$$

Le résidu  $r_{ij}$  correspond à l'erreur résiduelle  $\varepsilon_{ui}$  ou  $\varepsilon_{vi}$  - suivant l'espace de représentation considéré - associée à la modélisation du  $i^{\text{ème}}$   
 30 échantillon par la droite de régression de paramètres  $E_j$ . La solution à ce problème de minimisation non-linéaire nécessite une recherche de la droite définie par  $E_j$  parmi toutes les droites possibles. Afin de restreindre les calculs, on limite la recherche à un ensemble fini de  $p$  droites de régression, définies par  $p$  couples de points tirés aléatoirement parmi les échantillons de  
 35 l'espace de représentation étudié. Pour chacune des  $p$  droites, on calcule et

on trie les carrés des résidus de façon à identifier le carré du résidu carré qui présente la valeur médiane. La droite de régression est estimée comme celle qui fournit la plus petite de ces valeurs médianes des carrés des résidus.

Le fait de sélectionner la droite de régression sur l'unique carré  
5 du résidu médian, plutôt que sur l'ensemble des résidus, confère au processus de régression son caractère robuste. En effet, il permet de ne pas prendre en compte des résidus de valeurs extrêmes, susceptibles de correspondre à des points de données aberrants et donc de fausser la régression.

10 En testant par exemple  $p = 12$  droites, la probabilité qu'au moins un des  $p$  couples soit constitué de deux échantillons non-aberrants, c'est-à-dire représentatifs du mouvement dominant, est très proche de 1. Si la proportion d'échantillons aberrants est inférieure à 50%, ce que l'on suppose, un tel couple ne comportant pas d'échantillon aberrant fournit une droite de  
15 régression en meilleure adéquation avec le nuage d'échantillons – donc présentant un résidu carré médian plus faible – que n'importe quel couple de points comportant au moins un échantillon aberrant. On est alors quasiment assuré que la droite de régression finalement obtenue est définie par deux échantillons non aberrants, ce qui garantit la robustesse de la méthode vis-à-  
20 vis des échantillons aberrants.

Les droites de régression obtenues par estimation robuste dans chaque espace de représentation sont ensuite utilisées pour identifier les échantillons aberrants. Dans ce but, on calcule, en fonction de la valeur médiane du carré du résidu correspondant à la meilleure droite de régression  
25 trouvée, une estimation robuste  $\hat{\sigma}$  de l'écart-type des résidus associés aux échantillons non aberrants, sous l'hypothèse qu'ils suivent une distribution gaussienne, et on étiquette comme échantillon aberrant tout échantillon dont la valeur absolue du résidu excède  $K$  fois  $\hat{\sigma}$ . On peut avantageusement fixer la valeur de  $K$  à 2,5.

30 Toujours dans cette étape 3, on effectue enfin des régressions linéaires classiques, non robustes, sur les échantillons de chaque espace de représentation, en excluant les échantillons identifiés comme aberrants. Ces régressions fournissent des estimations affinées des paramètres  $(a_0, b_0)$  et  $(a_1, b_1)$  qui seront utilisées dans la suite du procédé.

35 L'étape suivante 4 effectue un test de linéarité des droites de régression dans chacun des espaces de représentation. Ce test a pour but

de vérifier que les nuages de points dans chaque espace sont effectivement approximativement distribués suivant des droites, ce que ne garantit aucunement l'existence – systématique – d'une droite de régression.

Le test de linéarité est effectué, dans chaque espace de représentation, en comparant à un seuil prédéterminé l'écart type des résidus issus de la régression linéaire portant sur les échantillons non aberrants. La valeur du seuil dépend de la normalisation temporelle appliquée aux vecteurs de mouvement à l'étape 1 du procédé. Dans le cas où, après normalisation, chaque vecteur représente un déplacement correspondant à l'intervalle de temps séparant deux trames entrelacées, soit 40 ms pour une transmission à 50 Hz, ce seuil peut avantageusement être fixé à 6.

Si au moins un des tests de linéarité effectués dans les deux espaces de représentation échoue, on considère que le champ de mouvement correspondant à l'image courante ne permet pas une estimation fiable d'un modèle de mouvement dominant. On positionne alors un drapeau signalant l'échec du processus d'estimation du mouvement dominant, et on traite l'image suivante.

Dans le cas contraire, on passe à l'étape suivante 5, qui consiste à vérifier que les pentes  $a_0$  et  $a_1$ , qui fournissent une double estimation du paramètre de divergence  $k$  du modèle de mouvement, ne diffèrent pas de manière significative. Le test d'égalité de deux pentes de régression est un problème connu, qui est traité dans certains ouvrages de statistique ; on pourra par exemple consulter le chapitre consacré à l'analyse de variance dans le livre de C.R. Rao « Linear Statistical Inference and its Applications » paru aux éditions Wiley (2<sup>ème</sup> édition). On effectue de manière classique ce test en calculant une pente de régression globale portant sur l'ensemble des échantillons non aberrants des deux espaces de représentation du champ de vecteurs de mouvement. On forme ensuite le rapport de la somme des carrés des résidus relative à cette estimation de pente globale sur l'ensemble des données, à la somme sur les deux espaces des sommes des carrés des résidus relatives aux régressions séparées – ne portant que sur les échantillons non aberrants. On compare ce rapport à un seuil prédéterminé ; si le rapport est supérieur au seuil, l'hypothèse d'égalité des pentes de régression dans les deux espaces de représentation du mouvement n'est statistiquement pas valide. On positionne alors un drapeau

signalant l'échec du processus d'estimation du mouvement dominant, et on traite l'image suivante. Dans le cas où le résultat du test est positif, on estime la valeur du coefficient de divergence  $k$  du modèle de mouvement dominant par la moyenne arithmétique des pentes de régression  $a_0$  et  $a_1$  obtenues dans chacun des espaces de représentation. Les paramètres  $t_x$  et  $t_y$  sont estimés respectivement par les valeurs des intercepts  $b_0$  et  $b_1$  issus des régressions linéaires dans les espaces de représentation.

Dans le cas où le modèle de mouvement est considéré comme valide, c'est-à-dire si les tests effectués dans les étapes 4 et 5 ont été passés avec succès, une classification du mouvement dominant est effectuée lors de l'étape suivante référencée 6.

Le vecteur  $\theta = (k, t_x, t_y)^t$  des paramètres estimés est exploité pour décider de la catégorie dans laquelle classer le mouvement dominant, à savoir :

- statique,
- translation pure,
- zoom pur,
- translation combinée avec un zoom.

L'algorithme de classification est basé sur des tests de nullité des paramètres du modèle, conformément au tableau ci-dessous :

Modèle	Paramètres		
Statique	$k = 0$	$t_x = 0$	$t_y = 0$
Translation	$k = 0$	$(t_x, t_y) \neq (0, 0)$	
Zoom	$k \neq 0$	$t_x = 0$	$t_y = 0$
Zoom + translation	$k \neq 0$	$(t_x, t_y) \neq (0, 0)$	

Selon une technique simple, les tests de nullité des estimations des paramètres du modèles peuvent être effectués par simple comparaison à un seuil de leur valeur absolue. Des techniques plus élaborées, basées sur une modélisation statistique de la distribution des données, peuvent également être employées. Dans ce cadre statistique, un exemple d'algorithme de décision de la nullité des paramètres du modèle basé sur des tests de vraisemblance est présenté dans l'article de P. Bouthemy, M. Gelgon et F. Ganansia intitulé « A unified approach to shot change detection and camera motion characterization », paru dans la revue IEEE Circuits and

Systems for Video Technology volume 9 n° 7, Octobre 1999, pages 1030 à 1044.

Une application de l'invention concerne l'indexation vidéo à partir de la sélection d'images clés.

5           En effet, le processus d'indexation vidéo débute généralement par un pré-traitement, qui vise à restreindre le volume des informations à traiter dans le flux vidéo à un ensemble d'images clés sélectionnées dans la séquence. Les traitements d'indexation de la vidéo, et en particulier l'extraction des attributs visuels, s'effectuent exclusivement sur ces images  
10 clés, dont chacune est représentative du contenu d'un segment de la vidéo. Idéalement, l'ensemble des images clés doit former un résumé exhaustif de la vidéo, et les redondances entre le contenu visuel des images clés doivent être évitées, de façon à minimiser la charge de calcul du processus d'indexation. Le procédé d'estimation du mouvement dominant à l'intérieur  
15 de chaque plan vidéo permet d'optimiser la sélection des images clés, à l'intérieur de chaque plan, vis-à-vis de ces critères, en l'adaptant au mouvement dominant. Il est par exemple possible de cumuler les translations horizontales (respectivement verticales) de l'image, estimées par le paramètre  $t_x$  (respectivement  $t_y$ ) à l'intérieur d'un plan, et d'échantillonner une  
20 nouvelle image clé dès que le cumul dépasse la largeur (respectivement la hauteur) d'une image.

Le procédé décrit peut également être exploité pour la génération de méta-données. Les mouvements dominants coïncident souvent avec les mouvements de caméra lors du tournage de la vidéo.  
25 Certains réalisateurs utilisent des séquences de mouvements de caméra particulières afin de communiquer au spectateur certaines émotions ou sensations. Le procédé décrit dans l'invention peut permettre de détecter ces séquences particulières dans la vidéo, et par conséquent de fournir des méta-données relatives à l'atmosphère créée par le réalisateur dans certains  
30 passages de la vidéo.

Une autre application de la détection de mouvement dominant est la détection ou l'aide à la détection de ruptures de plans. En effet, un changement abrupt des propriétés du mouvement dominant dans une séquence ne peut être engendré que par une rupture de plan.

35           Enfin, le procédé décrit dans l'invention permet l'identification, dans chaque image, du support du mouvement dominant. Ce support

coïncide en effet avec l'ensemble des pixels dont le vecteur associé n'a pas été identifié comme aberrant, au sens du mouvement dominant. La connaissance du support du mouvement dominant fournit une segmentation de l'objet qui suit ce mouvement. Cette segmentation peut être exploitée soit

5 pour effectuer une indexation séparée des objets constituant l'image, ce qui permet le traitement de requêtes partielles portant sur les objets et non sur la totalité des images, soit dans le cadre d'algorithmes de compression vidéo basés sur des objets, tels que par exemple ceux spécifiés dans la norme de compression vidéo MPEG-4.



## REVENDICATIONS

- 1 Procédé d'estimation d'un mouvement dominant dans une  
5 séquence d'images effectuant un calcul (1) d'un champ de vecteurs de mouvement associé à une image, définissant, pour un élément d'image de coordonnées  $x_i$ ,  $y_i$ , un ou plusieurs vecteurs de mouvement de composantes  $u_i$ ,  $v_i$ , caractérisé en ce qu'il effectue également les étapes suivantes:
- modélisation du mouvement (2) à partir d'une représentation  
10 paramétrique simplifiée :  
$$u_i = t_x + k.x_i$$
$$v_i = t_y + k.y_i$$
  
avec  
 $t_x$ ,  $t_y$  composantes d'un vecteur représentant la composante de  
15 translation du mouvement,  
 $k$  facteur de divergence caractérisant la composante de zoom du mouvement,  
- régression linéaire robuste (3) dans chacun des deux espaces de représentation du mouvement définis par les plans  $(x,u)$  et  $(y,v)$ ,  $x$ ,  $y$ ,  $u$  et  
20  $v$  représentant respectivement les axes des variables  $x_i$ ,  $y_i$ ,  $u_i$  et  $v_i$ , pour donner des droites de régression,  
- calcul des paramètres  $t_x$ ,  $t_y$ , et  $k$  (4, 5) à partir des ordonnées à l'origine et pentes des droites de régression.
- 25 2 Procédé selon la revendication 1, caractérisé en ce que la régression robuste (3) est la méthode de la moindre médiane des carrés qui consiste à chercher, parmi un ensemble de droites  $j$ ,  $r_{i,j}$  étant le résidu du  $i$ ème échantillon de coordonnées  $x_i$ ,  $u_i$  ou  $y_i$ ,  $v_i$ , par rapport à une droite  $j$ , celle fournissant la valeur médiane de l'ensemble des carrés des résidus qui  
30 est minimale.
- 3 Procédé selon la revendication 2, caractérisé en ce que la recherche de la moindre médiane des carrés des résidus (3) est appliquée  
35 sur un nombre prédéfini de droites déterminées chacune par un couple

d'échantillons tirés aléatoirement dans l'espace de représentation du mouvement considéré.

3 Procédé selon la revendication 1, caractérisé en ce qu'il effectue, après la régression linéaire robuste (3), une deuxième régression linéaire non-robuste permettant d'affiner les estimations des paramètres du modèle de mouvement.

4 Procédé selon la revendication 3, caractérisé en ce que la deuxième régression linéaire exclut les points dans les espaces de représentation dont le résidu de régression issu de la première régression robuste excède un seuil prédéterminé.

5 Procédé selon la revendication 1, caractérisé en ce qu'il effectue un test d'égalité (5) des coefficients directeurs des droites de régression calculées dans chacun des espaces de représentation (4), basé sur une comparaison des sommes des carrés des résidus obtenus premièrement en effectuant deux régressions séparées dans chaque espace de représentation, deuxièmement en effectuant une régression de pente globale sur l'ensemble des échantillons des deux espaces de représentation, et, dans le cas où le test est positif, qu'il estime le paramètre  $k$  du modèle par la moyenne arithmétique des coefficients directeurs des droites de régression obtenues dans chaque espace de représentation.

6 Procédé selon la revendication 1, caractérisé en ce que le mouvement dominant est classé dans une des catégories : translation, zoom, combinaison d'une translation et d'un zoom, image statique, selon les valeurs de  $t_x$ ,  $t_y$  et  $k$ .

7 Procédé selon la revendication 1, caractérisé en ce que le champ de vecteurs de mouvement est issu de l'encodage de la séquence vidéo considérée par un algorithme de compression utilisant la compensation du mouvement, tel que les algorithmes conformes aux normes de compression MPEG-1, MPEG-2 ou MPEG-4.

35

8 Application du procédé selon la revendication 1 à la sélection d'images clés, une image étant sélectionnée en fonction du cumul, sur plusieurs images, des informations relatives aux paramètres calculés  $t_x$ ,  $t_y$ , ou  $k$ .

5

9 Dispositif d'estimation d'un mouvement dominant dans une séquence d'images comportant un circuit de calcul (1) d'un champ de vecteurs de mouvement associé à une image, définissant, pour un élément d'image de coordonnées  $x_i$ ,  $y_i$ , un ou plusieurs vecteurs de mouvement de composantes  $u_i$ ,  $v_i$ , caractérisé en ce qu'il comporte également des moyens de calcul pour effectuer:

10

- une modélisation du mouvement (2) à partir d'une représentation paramétrique simplifiée :

15

$$u_i = t_x + k.x_i$$

$$v_i = t_y + k.y_i$$

avec

$t_x$ ,  $t_y$  composantes d'un vecteur représentant la composante de translation du mouvement,

20

$k$  facteur de divergence caractérisant la composante de zoom du mouvement,

- une régression linéaire robuste (3) dans chacun des deux espaces de représentation du mouvement définis par les plans  $(x,u)$  et  $(y,v)$ ,  $x$ ,  $y$ ,  $u$  et  $v$  représentant respectivement les axes des variables  $x_i$ ,  $y_i$ ,  $u_i$  et  $v_i$ , pour donner des droites de régression,

25

- un calcul des paramètres  $t_x$ ,  $t_y$ , et  $k$  (4, 5) à partir des ordonnées à l'origine et pentes des droites de régression.

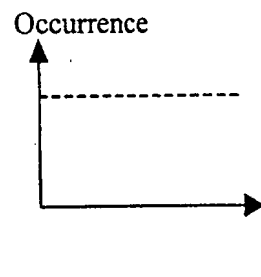
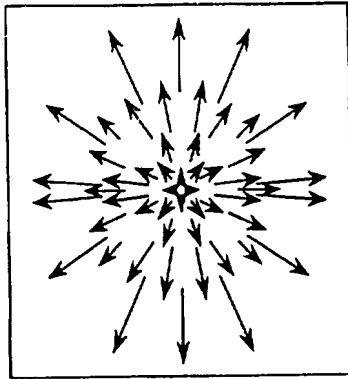
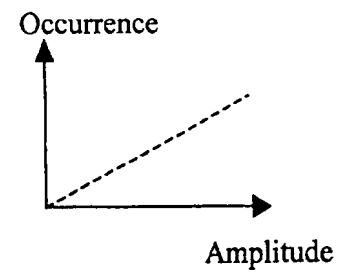
 $\theta$ 

FIG.1

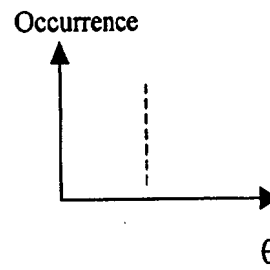
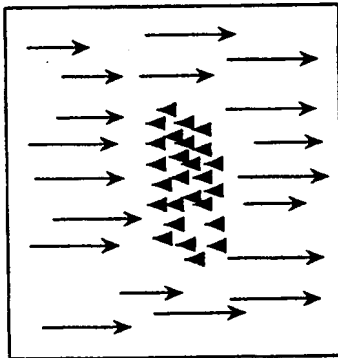
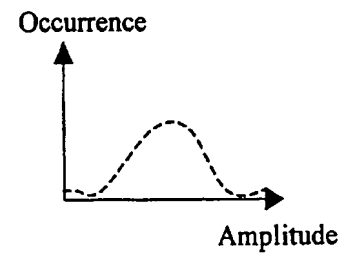
 $\theta$ 

FIG.2

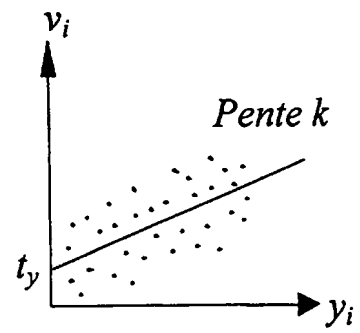
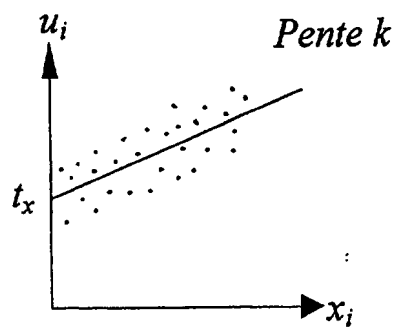


FIG.3

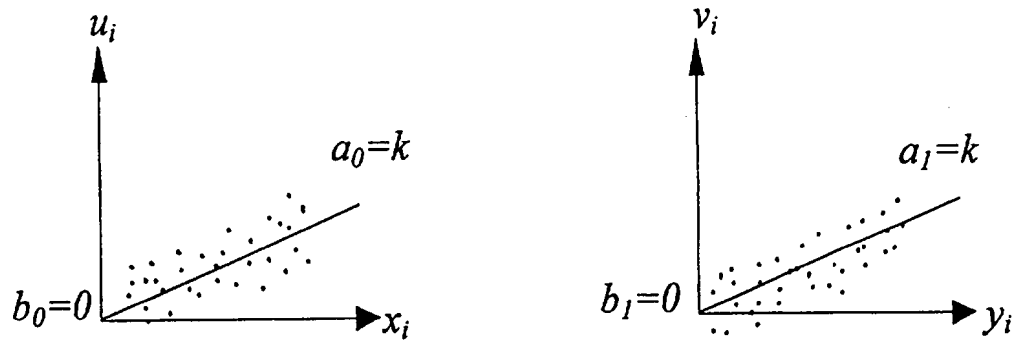


FIG. 4

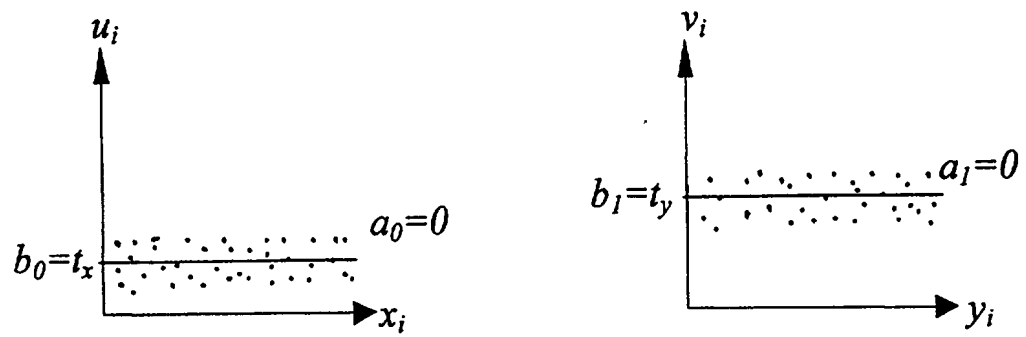


FIG. 5

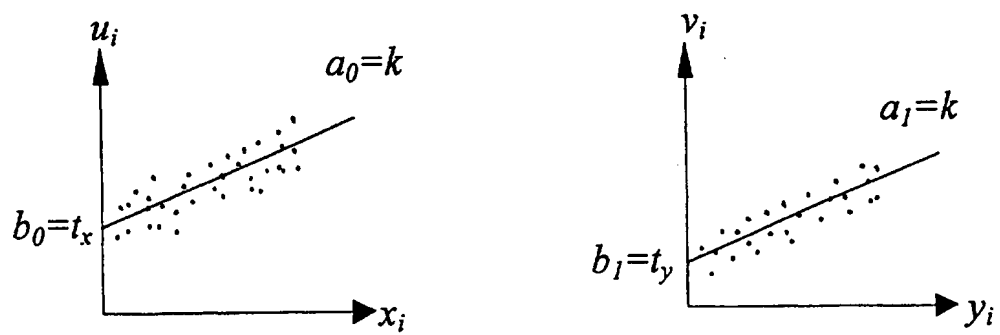


FIG. 6

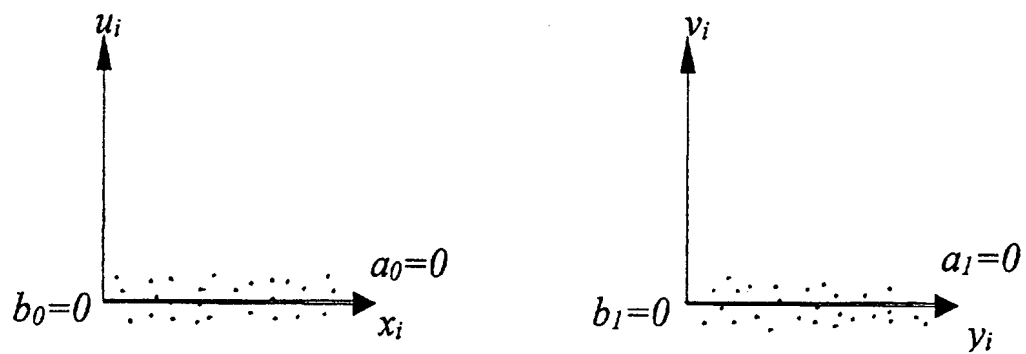


FIG.7

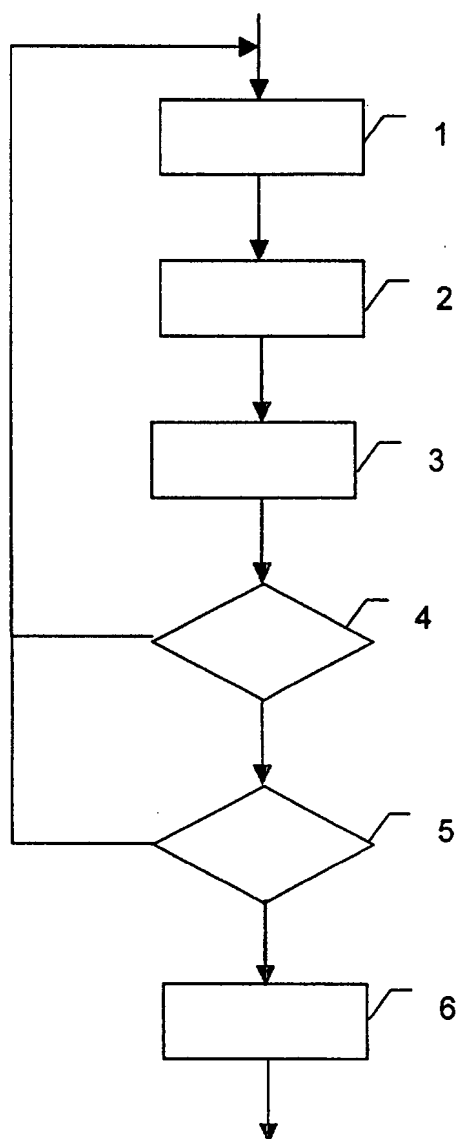


FIG.8

**RAPPORT DE RECHERCHE  
PRÉLIMINAIRE**

N° d'enregistrement  
national

établi sur la base des dernières revendications  
déposées avant le commencement de la recherche

FA 614864  
FR 0116466

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
A	<p>CLOUTIER L ET AL: "Segmentation and estimation of image motion by a robust method"</p> <p>PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING (ICIP) AUSTIN, NOV. 13 - 16, 1994, LOS ALAMITOS, IEEE COMP. SOC. PRESS, US, vol. 3 CONF. 1, 13 novembre 1994 (1994-11-13), pages 805-809, XP010146282 ISBN: 0-8186-6952-7</p> <p>* abrégé *</p> <p>* page 805, colonne de gauche, alinéa 2 *</p>	1-9	<p>H04N7/50 G06T7/20 H03M7/30</p>
A	<p>KUI ZHANG ET AL: "Global motion estimation and robust regression for video coding"</p> <p>ACOUSTICS, SPEECH AND SIGNAL PROCESSING, 1998. PROCEEDINGS OF THE 1998 IEEE INTERNATIONAL CONFERENCE ON SEATTLE, WA, USA 12-15 MAY 1998, NEW YORK, NY, USA, IEEE, US, 12 mai 1998 (1998-05-12), pages 2589-2592, XP010279424 ISBN: 0-7803-4428-6</p> <p>* page 2589, colonne de gauche, alinéa 2 *</p> <p>* page 2590, colonne de droite, alinéa 2 *</p>	1-9	<p><b>DOMAINES TECHNIQUES RECHERCHÉS (Int.CL.7)</b></p> <p>H04N G06T G06F</p>
A	<p>EP 1 050 849 A (THOMSON MULTIMEDIA SA) 8 novembre 2000 (2000-11-08)</p> <p>* le document en entier *</p> <p style="text-align: center;">-/--</p>	1-9	
Date d'achèvement de la recherche		Examineur	
3 septembre 2002		Gries, T	
<p><b>CATÉGORIE DES DOCUMENTS CITÉS</b></p> <p>X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire</p> <p>T : théorie ou principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons</p> <p>&amp; : membre de la même famille, document correspondant</p>			

**RAPPORT DE RECHERCHE  
PRÉLIMINAIRE**

N° d'enregistrement  
national

établi sur la base des dernières revendications  
déposées avant le commencement de la recherche

FA 614864  
FR 0116466

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
A	HIROHISA JOZAWA ET AL: "TWO-STAGE MOTION COMPENSATION USING ADAPTIVE GLOBAL MC AND LOCAL AFFINE MC" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE INC. NEW YORK, US, vol. 7, no. 1, 1 février 1997 (1997-02-01), pages 75-85, XP000678881 ISSN: 1051-8215 * abrégé * * page 75, colonne de droite, alinéa 3 *	1-9	DOMAINES TECHNIQUES RECHERCHÉS (Int.CL.7)
A,D	BOUTHEMY P ET AL: "A UNIFIED APPROACH TO SHOT CHANGE DETECTION AND CAMERA MOTION CHARACTERIZATION" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE INC. NEW YORK, US, vol. 9, no. 7, octobre 1999 (1999-10), pages 1030-1044, XP000853337 ISSN: 1051-8215 * le document en entier *	1-9	
A,D	GEREK O N ET AL: "Key frame selection from MPEG video data" VISUAL COMMUNICATIONS AND IMAGE PROCESSING '97, SAN JOSE, CA, USA, 12-14 FEB. 1997, vol. 3024, pt.2, pages 920-925, XP008007282 Proceedings of the SPIE - The International Society for Optical Engineering, 1997, SPIE-Int. Soc. Opt. Eng, USA ISSN: 0277-786X * figures *	1-9	
Date d'achèvement de la recherche		Examineur	
3 septembre 2002		Gries, T	
<p><b>CATÉGORIE DES DOCUMENTS CITÉS</b></p> <p>X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document Intercalaire</p> <p>T : théorie ou principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons &amp; : membre de la même famille, document correspondant</p>			



**RAPPORT DE RECHERCHE  
PRÉLIMINAIRE**

établi sur la base des dernières revendications  
déposées avant le commencement de la recherche

N° d'enregistrement  
national

FA 614864  
FR 0116466

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
A,D	<p>SMITH M A ET AL: "VIDEO SKIMMING AND CHARACTERIZATION THROUGH THE COMBINATION OF IMAGE AND LANGUAGE UNDERSTANDING TECHNIQUES"</p> <p>PROCEEDINGS OF THE 1997 IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. SAN JUAN, PUERTO RICO, JUNE 17 - 19, 1997, PROCEEDINGS OF THE IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, LOS ALAMIT,</p> <p>vol. CONF. 16, 17 juin 1997 (1997-06-17), pages 775-781, XP000776576</p> <p>ISBN: 0-7803-4236-4</p> <p>* le document en entier *</p>	1-9	
		DOMAINES TECHNIQUES RECHERCHÉS (Int.C.L.7)	
Date d'achèvement de la recherche		Examineur	
3 septembre 2002		Gries, T	
CATÉGORIE DES DOCUMENTS CITÉS			
<p>X : particulièrement pertinent à lui seul</p> <p>Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie</p> <p>A : arrière-plan technologique</p> <p>O : divulgation non-écrite</p> <p>P : document intercalaire</p> <p>T : théorie ou principe à la base de l'invention</p> <p>E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure.</p> <p>D : cité dans la demande</p> <p>L : cité pour d'autres raisons</p> <p>&amp; : membre de la même famille, document correspondant</p>			

**ANNEXE AU RAPPORT DE RECHERCHE PRÉLIMINAIRE  
RELATIF A LA DEMANDE DE BREVET FRANÇAIS NO. FR 0116466 FA 614864**

La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de recherche préliminaire visé ci-dessus.  
Les dits membres sont contenus au fichier informatique de l'Office européen des brevets à la date d'03-09-2002  
Les renseignements fournis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets, ni de l'Administration française

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
EP 1050849      A	08-11-2000	EP      1050850 A1	08-11-2000
		EP      1050849 A1	08-11-2000
		CN      1272747 A	08-11-2000
		JP      2000339475 A	08-12-2000
<hr/>			